

Contributions to the Theory of de Bruijn Cycles

André Alexander Campbell & Anant P. Godbole
 Department of Mathematics and Statistics
 East Tennessee State University

Bill Kay
 Department of Mathematics and Computer Science
 Emory University

April 11, 2013

Abstract

A de Bruijn cycle is a cyclic listing of length \mathcal{A} , of a collection of \mathcal{A} combinatorial objects, so that each object appears exactly once as a set of consecutive elements in the cycle. In this paper, we show the power of de Bruijn's original theorem, namely that the cycles bearing his name exist for n -letter words on a k -letter alphabet for all values of k, n , to prove that we can create de Bruijn cycles for the assignment of elements of $[n] = \{1, 2, \dots, n\}$ to the sets in any labeled subposet of the Boolean lattice; de Bruijn's theorem corresponds to the case when the subposet in question consists of a single ground element. The landmark work of [3] extended the agenda of finding de Bruijn cycles to possibly the next most natural set of combinatorial objects, namely k -subsets of $[n]$. In this area, important contributions have been those of [6] and [9]. Here we follow the direction of [1], who proved that, in a suitable encoding, de Bruijn cycles can be created for the subsets of $[n]$ of size in the interval $[s, t]; 0 \leq s < t \leq n$. In this paper we generalize this result to exhibit de Bruijn cycles for words with weight between s and t , where these parameters are suitably restricted.

1 Introduction

A de Bruijn cycle is a cyclic listing of length \mathcal{A} , of a collection of \mathcal{A} combinatorial objects, so that each object appears exactly once as a set of consecutive elements in the cycle. For example, the cyclic list 11101000 encodes each of the eight binary three letter words so that each appears exactly once. Viewed differently, and using a different coding, we see that the same list encodes all the 8 subsets of $[3] = \{1, 2, 3\}$, with the convention that, e.g., the string 101 represents the subset $\{1, 3\}$ – whose characteristic vector it happens to be. de Bruijn’s theorem states that this example is not an anomaly, and that de Bruijn cycles exist for the collection of n -letter words on the k -letter alphabet $\{0, 1, \dots, k - 1\}$, and thus, using a different coding, for all the multisets of $[n] = \{1, 2, \dots, n\}$ where each element may appear at most $k - 1$ times in a multiset. *A de Bruijn cycle is most often called a Universal Cycle (or U-cycle) in general contexts, but we shall not do so in this paper.*

Consider the poset $\mathcal{B} = \mathcal{B}_n$, the so-called Boolean Lattice of all subsets of $[n]$ ordered by inclusion. A *subposet* of \mathcal{B} is defined in this paper to be a collection of subsets of \mathcal{B} with the same Hasse diagram that is, moreover, consistent with the Hasse diagram of \mathcal{B} . For example, the W-poset or a 4-chain are each consistent with \mathcal{B}_3 but not with \mathcal{B}_2 . In Section 2 of this paper, we show the versatility of de Bruijn’s original theorem by showing that we can create de Bruijn cycles for the assignment of elements of $[n] = \{1, 2, \dots, n\}$ to the sets in *any labeled* subposet of the Boolean lattice; de Bruijn’s theorem corresponds to the case when the subposet in question consists of a single ground element. This result is equivalent to showing that de Bruijn cycles exist for the possible ways of filling in the elements of a collection of sets, labeled A, B, C , etc., whose Venn Diagram has a shape with specified intersection patterns.

The landmark work of [3] extended the agenda of finding de Bruijn cycles to possibly the next most natural set of combinatorial objects, namely k -subsets of $[n]$. Here the authors conjectured that de Bruijn cycles for the k -subsets of $[n]$ existed if and only if n was sufficiently large and $n \mid \binom{n}{k}$, which is an obvious necessary condition for the existence of the cycle in the natural coding (where, e.g., the set $\{a, b, c\}$ is encoded in the string by some permutation of its elements) since in this case each element of the set must appear equally often. In this area, important contributions have

been those of [6] and [9]. In [6], Hurlbert makes significant progress for the case of $k = 3, 4, 6$; the case of $k = 4$ is reduced to verification of two ground cases in [9]. Here we follow the direction of [1], who proved that, in a suitable encoding, de Bruijn cycles can be created for the subsets of $[n]$ of size in the interval $[s, t]; 0 \leq s < t \leq n$. For example, the string 1110011010 is a de Bruijn cycle of all the 2- and 3-subsets of $[4]$ if one uses a sliding window of length 4 and the characteristic vector coding for subsets. Once again, we remind the reader that setting $s = 0$ and $t = n$ and using a full (length n) encoding window for a subset yields de Bruijn's theorem! Thus the result in [1] can be viewed as a restricted version of de Bruijn's theorem. In this paper we generalize this result to exhibit de Bruijn cycles for words of length n over a k -ary alphabet and having weight between s and t , where the parameters s and t are suitably restricted. Viewed differently, this proves existence of de Bruijn cycles for multisets having between s and t elements, and where each element of $[n]$ may appear at most $k - 1$ times in the multiset. Setting $k = 2$ enables us to retrieve the result in [1] on subsets of cardinality between s and t . Results for de Bruijn cycles of multisets using the "classical" coding may be found in [8].

A good survey of some recent and not-so-recent papers on de Bruijn cycles may be found in [7]. See also [2], where de Bruijn cycles are proved to exist for graphs, hypergraphs, and such – using the simple and far-reaching device of invoking de Bruijn's theorem, where it all began, together with using a suitable encoding. And, last but not least, the reader is urged to study the very recent [5] and order it for his/her library.

2 Labeled Subposets of the Boolean Lattice, a.k.a. Venn Diagrams with Specified Intersection Patterns

A poset (P, \leq) is a set P with a relation \leq on P that is reflexive, transitive, and antisymmetric. The Hasse diagram of a poset is a representation of (P, \leq) using a pictorial device that represents its *transitive reduction*. Specifically, an "upward edge" is drawn from x to y if $x < y$ and there is no z such that $x < z < y$. The main result of this section is the following:

Theorem 1. *Let (P, \leq) be a fixed poset with elements of P being subsets of $[n]$ and $A \leq B$ if $A \subseteq B$. Assume furthermore that the Hasse diagram of P is consistent with that of the Boolean Lattice \mathcal{B}_n . There are then α^n assignments of the elements of $[n]$ to the elements of P , where α is the number of antichains in P . Furthermore, there exists a de Bruijn cycle of these assignments.*

Proof. We start with the first part of the result. Given any element $j \in [n]$ it is clear that if $j \in A$ for some $A \in P$, we must have $j \in B$ for each $B \geq A$, since \leq is the inclusion relation. It thus follows that the number of ways to assign element j to the sets in P is equal to the number of ways of labeling the elements of P with zeros and ones so that if A is labeled by a 1, then so is each $B \supseteq A$; this includes the case where all the elements of P are labeled with zeros. We call a coloring $c : P \rightarrow \{0, 1\}$ *unitarily up-closed* if $c(A) = 1 \Rightarrow c(B) = 1$ for each $B \in P$ with $B \geq A$. We will prove below that the set of unitarily up-closed colorings is in bijection with the antichains of (P, \leq) ; note that \emptyset is considered to be an antichain in P , as are all one-element sets:

For one direction, let c be a unitarily up-closed coloring on (P, \leq) . Let $M_c := \{A \in P : c(A) = 1 \text{ and } c(B) = 0 \text{ for every } B \leq A\}$ i.e., the set of elements of P minimal with respect to receiving the value 1 under c . Clearly, M_c forms an antichain, since if there exist $A, B \in M_c$ with $A \leq B$, then A does not meet the constructive criteria of M_c . This provides a map from colorings to antichains. To see that the map is injective, let c and d be two different unitarily up-closed colorings, and suppose that $M_c = M_d$. Fix $A \in P$. Suppose that $c(A) = 1$. If A is minimal, then we have that $d(A) = 1$ since $M_c = M_d$. If A is not minimal, it is above some $B \in M_c$. $B \in M_d$ by hypothesis, and since $B \leq A$, $d(A) = 1$ as d is unitarily up-closed. Now suppose that $c(A) = 0$. Since A is not above *any* element of M_c , A is also not above any element of M_d . Hence $d(A) = 0$, as if $d(A) = 1$ we would have that A was above some minimal element under d .

To see the reverse, notice that for a fixed antichain \mathcal{A} , we can define a unitarily closed up-coloring $c = c_{\mathcal{A}}$ by the rule $c(A) = 1$ iff $B \leq A$ for some $B \in \mathcal{A}$, 0 otherwise. We claim that $M_c = \mathcal{A}$, and hence this provides a $1 - 1$ inverse to our $1 - 1$ function. Since no element below any element of \mathcal{A} receives color 1, clearly $\mathcal{A} \subseteq M_c$. On the other hand, every element A with $c(A) = 1$ is above some element of \mathcal{A} , and since M_c contains every element

of \mathcal{A} , no other element receiving color 1 can belong to M_c , and we are done.

It follows that the “fate” of each element of $[n]$, i.e., which sets in P it belongs to, can be determined in α ways, and thus there are α^n assignments of the elements of $[n]$ to the sets in P .

To prove the second part, we note that each of the α ways of assigning element j to the sets in P can be coded using letters from an alphabet Λ of size α . By de Bruijn’s theorem, there exists a cycle of the elements of $\Lambda^{[n]}$, and thus any such de Bruijn cycle may be viewed as a de Bruijn cycle of the assignment of elements of $[n]$ to the sets in P . This completes the proof; specific examples follow. \square

Examples. The two letter words on a three letter alphabet $\{a, b, c\}$ may be cycled as follows:

$$cc \rightarrow ca \rightarrow aa \rightarrow ab \rightarrow bb \rightarrow bc \rightarrow cb \rightarrow ba \rightarrow ac.$$

We let the letters of the alphabet code for the “allowable configurations” corresponding to the three antichains of the 2-chain with Hasse diagram

$$\begin{array}{c} B \\ | \\ A \end{array}$$

This is done by using a coding such as

$$a = \begin{array}{c} 0 \\ 0 \end{array}; \quad b = \begin{array}{c} 1 \\ 0 \end{array}; \quad c = \begin{array}{c} 1 \\ 1 \end{array},$$

where, e.g., $\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}$ indicates that $j \in B; j \notin A$. Thus the de Bruijn cycle $ccaabbcb$ above may be written as

$$\begin{array}{cccccccc} 1 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{array}$$

To decode the cycle, we read the above stack from left to right in sliding groups of two columns at a time, the first and second columns describing to which sets elements 1 and 2 belong. Since $n = 2$, this is all we need. The actual de Bruijn cycle of assignments of $\{1, 2\}$ to the two sets in the poset is then as follows:

$$\begin{array}{cccccccc}
\{1, 2\} & \{1\} & \emptyset & \{2\} & \{1, 2\} & \{1, 2\} & \{1, 2\} & \{1\} & \{2\} \\
| & | & | & | & | & | & | & | & | \\
\{1, 2\} & \{1\} & \emptyset & \emptyset & \emptyset & \{2\} & \{1\} & \emptyset & \{2\}
\end{array}$$

The same may be done for a poset with any Hasse diagram. If, e.g., we have the W-poset with 13 antichains, then

(i) We start with a de Bruijn cycle for all the n -letter words on an alphabet such as $\{A, B, \dots, M\}$.

(ii) We then rewrite this as an assignment of the elements of $\{1, 2, \dots, n\}$ to the five sets in the poset, using a sliding stack of height 5 and width n .

(iii) Finally, we may “draw” the 13^n realizations of the poset, i.e., the assignments of n elements to 13 sets in a Venn diagram.

Note, however, that going from steps (i) to (ii) to (iii) leads to a sequential loss of data compression, and thus (i) gives the recommended de Bruijn cycle.

3 Words with Weights in Prescribed Intervals

The main result of this section is the following:

Theorem 2. *Let $k, n \in \mathbb{Z}^+$. Consider n letter words $w = (w_1, w_2, \dots, w_n)$ on the k -letter alphabet $\Lambda = \{0, 1, \dots, k-1\}$, and define the weight $h(w)$ of w by $h(w) = \sum_{i=1}^n w_i$. Let $s, t \in \mathbb{Z}^+$ satisfy $0 \leq s < s+k-1 \leq t \leq n(k-1)$. Let \mathcal{W} be the collection of all such words with weights between s and t . Then there exists a de Bruijn cycle of the elements of \mathcal{W} .*

Proof. We create a digraph $D = (V, E)$ with vertex set equal to the set of all $n-1$ -letter words over the alphabet Λ and having weights between $s-(k-1)$ and t . A directed edge is drawn from vertex v_1 to vertex v_2 if (i) the last $n-2$ letters of v_1 are the same as the first $n-2$ letters of v_2 and if (ii) the edge labeled by the concatenation of the letters in v_1 and v_2 yields an n -letter word with weight between s and t .

It is clear that the indegree $i(v)$ and outdegree $o(v)$ of any vertex v are both equal to the number of prefixes (or suffixes) that yield an edge label of legal weight; if, e.g., $h(v) = s - (k - 1)$ or $h(v) = t$, then $i(v) = o(v) = 1$. We will next show that D has a single weakly connected component; this will show that it is Eulerian [10], and the Eulerian path will give the required de Bruijn cycle.

Weak connectedness will be established by showing that there exists a path between any vertex in D and a special sink vertex SV of weight s that consists of a uniquely determined number $a \geq 1$ of letters x followed by $n - 1 - a \geq 0$ letters that are all $x + 1$. The rest of the proof of Theorem 2 consists of demonstrating four embedded lemmas, that state that we can (i) connect any vertex of weight $\geq s + 1$ to one of weight s ; (ii) connect any vertex of weight $\leq s - 1$ to one of weight s ; (iii) connect any vertex of weight s to one with weight s and letters $x, x + 1$; and (iv) connect any vertex with the right number of x s and $x + 1$ s to SV .

Lemma 3. *Let v have weight $h(v)$ in the range $[s + 1, t]$. Then there is a path from v to another vertex of weight s .*

Proof. We first cycle any zeros in the front of the vertex to the end. These steps may be taken without changing the weight of the vertex, and so that the weight of the words representing the edges is also maintained at $h(v)$. The resulting vertex, v' , has weight $h(v') = h(v)$ that may be as high as t , and has first letter different from zero. If $h(v') = t$, then the only allowable letter we can add, when the first letter is dropped, is zero. This maintains the edge weight at t , while leading to the new vertex weight $h(v'')$ satisfying the conditions $t > h(v'') \geq t - (k - 1) \geq s$. If $h(v') = t - r \geq s + 1$, we replace the first letter f by $\min\{f - 1, r\}$. This either reduces the vertex weight by one or changes it to $t - r - f + r = t - f \geq t - (k - 1) \geq s$. The edge weight might increase from $t - r$ to t , but this is OK. We now see that v'' either has weight s or, if not, has weight smaller than $h(v')$. We repeat the above process until we reach a vertex with weight s . \square

Lemma 4. *Let v have weight $h(v)$ in the range $[s - (k - 1), s - 1]$. Then there is a path from v to another vertex of weight s .*

Proof. If $h(v) = s - (k - 1)$, we drop the first letter and add the letter $k - 1$. This makes the edge label equal to s , and the vertex weight does not decrease.

If the vertex label has increased, we have made progress towards our target. If it has not, due to the first letter in v being $k - 1$, we repeat the process till we get to a vertex with first letter smaller than $k - 1$. If the starting weight is $h(v) = s - r$; $1 \leq r \leq k - 2$, we drop all letters $k - 1$ at the front of the word and append $k - 1$ s to the end until the first letter is smaller than $k - 1$, thus arriving at the vertex v' . We now add the letter $\max\{r, f + 1\}$, where f is the first letter of the vertex v' . This leads to the edge weight becoming either s (if $r \geq f + 1$) or $s - r + f + 1 \leq t$ (if $r \leq f + 1$). The vertex weight goes up by 1, if $f + 1 \geq r$ and becomes $s - f$ if $f + 1 \leq r$. We now iterate the above process until the vertex weight becomes s . \square

Lemma 5. *Let $v = (v_1, \dots, v_{n-1})$ have weight $h(v) = s$. Then there is a path from v to another vertex of weight s and consisting entirely of letters that are either x or $x + 1$ (and at least one letter x).*

Proof. Type A letters are those in the set $\{0, 1, \dots, x\}$, while type B letters belong to the set $\{x + 1, \dots, k - 1\}$. Let a and b be the desired number of x s and $x + 1$ s. Depending on how many type A and type B letters v has, we will need the path from v to decrease a certain number of type B letters to either x or $x + 1$, and to increase some or all of the type A letters to either x or $x + 1$. Start with v_1 . If v_1 is a type A letter we change it to either x or $x + 1$ as needed. This possibly increases the vertex and edge weights, and the condition $t \geq s + (k - 1)$ keeps both weights legal. (The weights don't have to increase since we may replace the x with an x , causing no change in vertex weight; or x might equal 0, in which case there is no change in edge weight.) If v_1 is of type B we again change it to either x or $x + 1$, causing a possible drop in vertex weight. We now repeat the process with the first letter v_2 of the new vertex v' *as long as the new vertex weight $h(v')$ satisfies $t - h(v') \geq x$ or $h(v') - (s - (k - 1)) \geq x$* . No edge traversal in the digraph is undertaken that leads to a vertex that is closer than x in weight from the two extreme vertex weights, namely $s - (k - 1)$ and t . If such a “dangerous” occurrence is imminent, we abort such a move, cycling instead until we have the opportunity to increase a “dangerously low weight” or decrease a “dangerously high vertex weight.” The process is repeated until we have the required numbers a, b of symbols x and $x + 1$ respectively. An example is given after the proof of Corollary 3. \square

Lemma 6. *Let v have weight $h(v) = s$ and be composed entirely of x s and $x + 1$ s as in Lemma 3. Then there is a path from v to the sink vertex*

$SV = (x, x, \dots, x, x+1, \dots, x+1)$ having the same number of x s and $x+1$ s as v .

Proof. The proof is similar to that of Lemma 5. We first identify both x s and $x+1$ s in v that are “out of place”. Clearly the number of out of place x s must equal the number of out of place $x+1$ s, and the latter all appear before the former. We cycle the letters of the word until we arrive at the first out of place $x+1$, delete it, and add an x at the end. This decreases the vertex weight by one. We continue cycling the word and replacing $x+1$ s by x s until the vertex weight is within x of the minimum legal weight, i.e. $s - (k-1)$. The next phase is to increase the vertex weight by replacing (cyclically) out of place x s by $x+1$ s, until the vertex weight is within x of the maximum allowable, i.e. t . We alternate this process until the two letters are all cyclically in the right places, and then cycle until we reach SV . \square

Lemmas 3, 4, 5, and 6 together complete the proof of Theorem 2. \square

We define a *Redundant de Bruijn Cycle* of a collection of \mathcal{A} combinatorial objects to be a cycle of length $\mathcal{A}' > \mathcal{A}$ so that each object appears exactly once as a set of consecutive elements in the cycle, and thus so that $\mathcal{A}' - \mathcal{A}$ consecutive elements are redundant objects of another type.

Corollary 7. *For each $t \geq k-1$, there exists a redundant de Bruijn cycle of the $\mathcal{A}(n, k, t)$ n -letter weight t words over $\{0, 1, \dots, k-1\}$ that is of length $\mathcal{A}'(n, k, t) = \mathcal{A}(n, k, t)(1 + o(1))$; k, t fixed, $n \rightarrow \infty$.*

Proof. We set $s = t - (k-1) \geq 0$ in Theorem 2, and obtain a de Bruijn cycle of all words with weight between $t - (k-1)$ and t . The length of this cycle is

$$\sum_{j=t-(k-1)}^t \mathcal{A}(n, k, j) = \mathcal{A}(n, k, t)(1 + o(1)),$$

since for fixed t and large n , $\mathcal{A}(n, k, j)$ is a monotone function of j ; $0 \leq j \leq t$ with $\mathcal{A}(n, k, j)/\mathcal{A}(n, k, j+1) \rightarrow 0$, $n \rightarrow \infty$. This completes the proof. \square

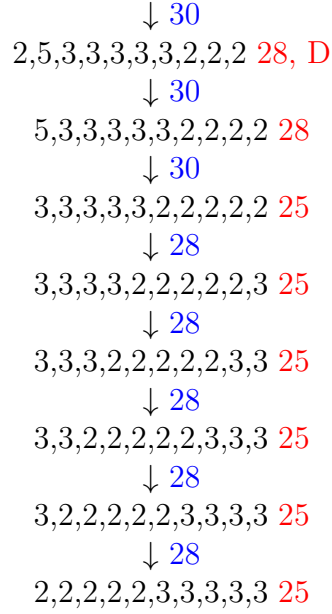
An Example: The Algorithm in Lemma 5 in Action. We choose v to be $(0, 0, 0, 2, 2, 5, 5, 5, 3, 3)$; $s = 25, t = 30, k = 6, n = 11$. The target vertex

is $SV = (2, 2, 2, 2, 2, 3, 3, 3, 3, 3)$. Legal vertex weights are between 20 and 30, and legal edge weights are between 25 and 30. We declare a dangerous situation to be one in which further progress would lead to vertex weights of 20, 21, 29, or 30 – unless we cycle to a vertex with appropriate first letter. Below, red numbers represent vertex weights, blue numbers represent edge weights, and the symbol D represents “Danger.” We proceed as follows:

```

0,0,0,2,2,5,5,5,3,3 25
      ↓ 28
0,0,2,2,5,5,5,3,3,3 28, D
      ↓ 28
0,2,2,5,5,5,3,3,3,0 28, D
      ↓ 28
2,2,5,5,5,3,3,3,0,0 28, D
      ↓ 30
2,5,5,5,3,3,3,0,0,2 28, D
      ↓ 30
5,5,5,3,3,3,0,0,2,2 28
      ↓ 30
5,5,3,3,3,0,0,2,2,2 25
      ↓ 27
5,3,3,3,0,0,2,2,2,2 22, D
      ↓ 27
3,3,3,0,0,2,2,2,2,5 22, D
      ↓ 25
3,3,0,0,2,2,2,2,5,3 22, D
      ↓ 25
3,0,0,2,2,2,2,5,3,3 22, D
      ↓ 25
0,0,2,2,2,2,5,3,3,3 22
      ↓ 25
0,2,2,2,2,5,3,3,3,3 25
      ↓ 28
2,2,2,2,5,3,3,3,3,3 28, D
      ↓ 30
2,2,2,5,3,3,3,3,3,2 28, D
      ↓ 30
2,2,5,3,3,3,3,3,2,2 28, D

```



thus arriving at the sink vertex.

4 Further Research

It would be interesting to improve Theorem 2 so that the range of the weight of the word is as small as possible, or else to prove that the range in Theorem 2 is the best possible. Also, one might ask how and to what extent one can show existence of de Bruijn cycles for unordered posets. Last but not least, can results be proved for (labelled as well as unlabelled) subposets of mother posets other than the Boolean Lattice?

5 Acknowledgments

The research of AG was supported by NSF Grant 1004624.

References

- [1] A. Blanca and A. Godbole (2011). “On Universal Cycles for new classes of combinatorial structures,” *SIAM J. Discrete Math.* **25**, 1832–1842.
- [2] G. Brockman, B. Kay, E. Snively (2010). “On universal cycles of labeled graphs,” *Electr. J. Combinatorics* **17**, Paper R4.
- [3] F. Chung, P. Diaconis, and R. Graham (1992). “Universal cycles for combinatorial structures,” *Discrete Math.* **110**, 43–59.
- [4] N. G. De Bruijn (1946). “A combinatorial problem,” *Nederl. Akad. Wetensch. Proc.* **49**, 758–764.
- [5] P. Diaconis and R. Graham (2011). *Magical Mathematics: The Mathematical Ideas that Animate Great Magic Tricks*, Princeton University Press.
- [6] G. Hurlbert (1994). “On universal cycles for k -subsets of an n -set,” *SIAM J. Discrete Math.* **7**, 598–604.
- [7] V. Horan and G. Hurlbert (2013+). “Universal Cycles for weak orders,” Preprint, see <http://arxiv.org/pdf/1203.5169.pdf>
- [8] G. Hurlbert, T. Johnson, and J. Zahl (2009). “On universal cycles for multisets,” *Discrete Math.* **309**, 5321–5327.
- [9] Y. Rudoy (2013+). “An inductive approach to constructing Universal Cycles on the k -subsets of $[n]$ ”, Preprint, see <http://arxiv.org/pdf/1209.4662v1.pdf>
- [10] D. West (1996). *Introduction to Graph Theory*, Prentice Hall, New Jersey.